UK England Chapter
2020 Cycle 1 Large - Beyond the Net Programme
Beyond the Net Large Grants Programme

Project Title: IPv6 Matrix Project Migration

Grant ID: G-202006-02895
Report Type: BNet Large Report - Interim
Due: October 15, 2021
Report ID: 886

español | français

Draft / Pending Submission **Received** Action Required Approved

▼ Table Of Contents

Due On: 10/15/2021

Submitted On: 10/16/2021

▼ Report Details

Reporting Period Start Date: 1/5/2021

Reporting Period End Date: 10/15/2021

Report Type: Interim

In your grant application you expected to complete the following activities by the end of your project. Please select the activities your project team undertook in this reporting period and the status of each activity in the form of a percentage complete.

Activities	Name	Description
#1	Audit of current IPv6 Matrix system	This first phase is a full audit of the two servers: 1. The Web server * its operating system * its software components including libraries and software scripts used * its database structure * its data repositories * its back-up systems * its regular operational and updating maintenance operations 2. the Crawler * its operating system * its software components, including libraries and software scripts used * its database structure * its data resositories * its back-up systems * its regular operational and updating maintenance operations
#2	Implementation of new Virtual Machine Environment	Scoping and design of new Virtual Machine Environment for the new IPv6 Matrix virtual machine * choice of operating system and its installation * memory, storage and processing requirements * network resources * back-up systems * potential for replication of resources
#3	Web Server Migration	Full migration of the Web server and its back-end database, from the Web Server hardware to the new Virtual Machine environment. Includes: - fixing and enhancing any security related bugs or weaknesses in the current Web server environment - fully documenting the new Web server environment - automating maintenance tasks such as updating the database, back-ups, rotating log files etc. - designing and implementing a way for third parties to have access to the raw data - packaging the Web server in a way that makes it easily replicable elsewhere in the world. The above must take place in a seamless manner - thus the existing IPv6 Matrix Web Site will switch from physical to virtual at a specific point in time during a limited maintenance window.

#4	Crawler Migration	Full migration of the Crawler and its database from the current Crawler hardware to the new Virtual Machine environment. Includes: - updating the Crawler to use modern libraries for its functions - updating the Crawler logic to reduce network load when tracerouting repeated addresses - fixing and enhancing any security related bugs or weaknesses in the current Crawler environment - fully documenting the new Crawler environment - automating maintenance tasks such as updating the database, back-ups, rotating log files etc. - updating the database of 1 Million most popular Web sites, taking into account the "no testing list" - packaging the Crawler in a way that makes it easily replicable elsewhere in the world. The above must take place in a seamless manner - thus the existing IPv6 Matrix Crawler will switch from physical to virtual at a specific point in time during a limited maintenance window. This is likely to happen between two successive crawls.
#5	IPv6 Matrix Updated Project Presentation	Presentation of the IPv6 Matrix migration Project in Webinars and/or conferences, describing lessons learnt, the project itself, as well as its latest results. The brief should also include a call for further third party hosts that could run the Crawler from their part of the world. This information is the main output of the project. Thirdly, we plan to issue a call for other organisations to further analyse and display the 500+Gb of data which we have collected since 2010, as a showcase of how the Internet is evolving and preparing itself for the future.

Activities Completed: Activity #1

#1 - Activity Percentage: 100%

#1 - Activity Narrative:

The first phase of the project was intended to be a full audit of the two servers - a webserver running CentOS 5, and a crawler running Ubuntu 4.4.
Unfortunately the servers suffered hardware failures before the project could be started, and so the audit had to be completed on a recovered hard disk image. These had been mounted to a virtual machine, however they had limited functionality. There was also a lot of retired functionality on the servers from an initial rewrite.
This first project involved, as much as possible, a full audit of the code on both disk images to determine aspects such as:
- how data is passed between the crawler and webserver;
- versions of software and libraries required for each system to function;
- a high level functionality of the two systems

This was undertaken fully so as to be ready to proceed with the activities following Activity 1.

Were the activities proposed and undertaken adequate and appropriate to address the overarching objective? Please describe.

A B-plan had to be devised for carrying out this work with:
1. building two temporary rigs for recovering all the lost data and simulating the two failed servers once the data was fully recovered
2. packaging all of the data into two full disk images from each of the failed servers. This 1+Tb of data was stored in Google Drive for easier transfer
3. building a development environment, importing each of the disk images and therefore having both Crawler and Web server on the same machine.

As a result, the rest of the project can now take place.

Project Area: Measuring the Internet.
Describe how your project supported the 2020 Project Area you originally selected in your funding application.

The reviving of the failed IPv6 Matrix served the aim of "measuring the Internet" so the grant arrived at exactly the right moment.

Describe how your project helped increase the presence of your Chapter/SIG within your community or region.

Not ready yet. Once Activity 2 and 3 will reach enough progress, a message will be sent to ISOC UK members. The first aim is to get the Web Server up and running again in a safe environment with existing data.

Did your project team collaborate with another Internet Society Chapter/SIG? No

What were the main challenges your project faced? How did you address them?

Total failure of the existing servers in a time of COVID was not foreseen.
Previous reports explain this in detail.

We did not give up and tried many solutions to get the servers to a contractor that knew what they were doing. The extra time allowed us to find someone competent and eager to get the project working.

What were the main successes of your project?

1. Getting the physical servers out of the university premises and to London
2. extracting the data from the failed servers and building 2 virtual test rigs
3. saving the data on Google Drive
4. successfully onboarding a contractor who managed to install his own test rig and development machine and start the actual work

Explain how your project provided your Chapter/SIG an opportunity to develop new and/or strengthen existing relationships and partnerships at the local level.

We managed to build a relationship with a competent contractor (Silvermouse) for UK Chapter technical projects.
We strengthened our relationship with Mythic Beasts regarding hosting of the project once it is ready to go live.

What were the main lessons you learned through implementing this project?

So far, two lessons were learned:
1. Murphy's Laws concerning the migration of the Servers
2. Don't ever give up

Following the completion of your project, what actions or activities do you plan to undertake (if any) to maintain relationships with the project participants?

Too early to decide.

Could other Chapter/SIGs learn from this project and possibly replicate it in their own communities? What advice would you provide to others Chapter/SIGs who might want to do so?

To be decided once the project will be complete. When it comes to Activity #1, the lesson to learn is "don't give up".

How could your Chapter/SIG build on this project moving forward?

Yes, ultimately, when the project is complete.

How many new Chapter/SIG members did your project attract *in this reporting period?* 0

Please select the measurements used during this reporting period to assess your project's impact and the number associated with each measurement:

Project Area: Measuring the Internet

KPI 1: Packaging and Full transfer of Web Server into a Virtual Machine environment. - Web Server needs to run as seamlessly and reliably as the current dedicated server. - Web Server needs to have as many of its currently manual maintenance tasks automated. - Web Server needs to be secured using current security practice. (https optional)

KPI 2: Optimisation, Packaging, Full transfer of Crawler into a Virtual Machine environment. - Crawler needs to run as seamlessly and reliably as the current dedicated server. - Crawler needs to have its crawling algorithms optimised to reduce redundancy and load on network and server - Crawler needs to have as many of its currently manual maintenance tasks automated. - Crawler needs to be secured using current security practice.

KPI 3: Safe Storage of currently collected data - Making this data more easily usable by third parties - Allowing third party analysis - Securing Data integrity for a minimum of ten more years - Presentation of Results in at least one or more major Internet event

KPI 4: Roll out of second Crawler from another world location. - Confirmation that packaging of the Crawler and Web Servers work

Choose which of the KPI's you're reporting on:	Key Performance Indicator #1, Key Performance Indicator #2
Key Performance Indicator 1:	50%
Key Performance Indicator 2:	30%

General Comments:

None of the KPIs have been achieved yet as we are still in early phase of the project, but the major hurdle of getting the servers replicated into a stable and safe environment is past.
The original server hardware (two large 1U rack servers, 1 HPDL360P and 1 HP DL140) have now been fully dismantled and disposed of by recycling - a special firm collected component parts, including the recycling of the rare metals on the motherboards. This case at no cost to the project, but we felt it needed to be done for the project to adhere to minimise its carbon footprint.

▼ Budget Information

REFERENCE: Budget from Grant Request ID #G-202006-02895

Category	Amount	Comment
Personnel		
Contractual	$22,500.00	This is the total value of the contract for professional contractors to undertake the project. The complexity of the task requires experienced contractors with a multiple set of skills: - Linux Admin - including three different types of Linux - the old types (CenTOS and an early version of Ubuntu) and the latest version of Ubuntu - Java / Jason - Python (in its various forms since 2010) - a vast array of technical skills mandated by the technologies used, as described in the 2014 project report Two attempts were made to make this a Group Design Project for students at Southampton University, but the complexity of the task requires a full time, experienced Team.
Equipment		
Travel	$2,000.00	In today's world, travel is kept at a minimum due to the worldwide Pandemic. However, it is expected that the project will benefit from at least two face to face meetings with both the representative from Mythic Beasts and the chosen Contractor in attendance. The figure provided is both a conservative figure and also optional, in case the pandemic does not allow for a face to face meeting. It is worth mentioning that the servers are currently physically in Southampton and the Mythic Beasts headquarters are in Cambridge. Travel might require physical movement of the servers.
Other Direct Costs	$200.00	This pertains to presentation material and consumables for the project from start to completion and includes general office sundries.
Total	**$24,700.00**	

Please upload your actual project expenses using the template found here.

Below, please add the high level amounts for each of the categories you are utilizing for this project.

How did you use Project Funds?	Other Direct Costs
Other Direct Costs Amount:	$1,000.00

Other Direct Costs Comment:

Too early to assess so far how much the initial challenged with the project affected the project costs. A cost overrun of $1000 has been set aside for the time being - physical transport of the servers, Hard Disk data recovery, temporary rig hosting, current monthly storage at Google Drive (ongoing but ending soon).
We are hoping to offset this added cost through reduction of Travel costs - but this is not yet confirmed.

Total Amount:	$1,000.00
During this reporting period, did you make any budget changes?	Yes
Was the change greater than 20% of your total grant award amount? :	No

▼ Documents

Please upload documents you believe will enhance your grantee report. These may include photo files, project deliverables (study findings, policy document, technical specifications), diagrams, etc.
Most file formats may be uploaded except for video and audio files.

REPORT ATTACHMENTS

BtN-Large-Financial-Information-Report_2021-10-16.xlsx
Financial Report Information
Added by Olivier Crepin-Leblond at 4:20 AM on October 16, 2021

elephant failed disk.jpeg
Other Document - Grantee Report | **Failed Disk with date from 2012**
Added by Olivier Crepin-Leblond at 3:46 AM on October 16, 2021

Crawler-ipv6matrix-org.jpeg
Other Document - Grantee Report | **Recycled Servers**
Added by Olivier Crepin-Leblond at 3:43 AM on October 16, 2021

elephant capacitors.jpeg
Other Document - Grantee Report | **Potential cause of hardware failure (blown capacitors)**
Added by Olivier Crepin-Leblond at 3:42 AM on October 16, 2021

Scan Silvermouse ISOC IPv6 Matrix Contract.pdf
Other Document - Grantee Report | **Contract with Silvermouse (contractor)**
Added by Olivier Crepin-Leblond at 3:42 AM on October 16, 2021

250120102466.jpg
Other Document - Grantee Report | **The two servers as they used to exist**
Added by Olivier Crepin-Leblond at 3:37 AM on October 16, 2021

Need Help? email us: fluxx_help@isocfoundation.org